評論建置與維護大型資料庫，以協助瞭解教育政策與措施執行成效以及規劃革新策略

彭森明
國立清華及中山大學榮譽講座

一、前言

當今資訊時代，各行業都重視大型資料庫之建置與運用，教育也不例外。先進國家如美國、英國、德國、澳洲等都有專職機構，主導全國性大型教育資料庫的永續規劃與建置工作。國際組織如聯合國的經濟合作與開發組織（Organisation for Economic Co-operation and Development，OECD）主導的 Program for International Student Achievement（PISA）以及 International Association for the Evaluation of Educational Achievement（IEA）主導的 Trends in International Mathematics and Sciece studie（TIMSS）和 Progress in International Reading Literacy Study（PIRLS）也引領會員國參與學生學習成果與相關因素的資料收集、分析、研究與評比。大型教育資料庫的建置與運用，已成為深受重視的國際趨勢。

臺灣近二十年來也開始重視到這項發展，推行一些國家級方案，如*臺灣*教育長期追蹤資料庫（Taiwan Education Panel Survey， TEPS），臺灣學生學習成就評量資料庫（TASA）以及高等教育資料整合計畫等。可惜的是，一些成型的方案，如高教資料整合系統，卻因不明原因被中斷了，連建置好的硬體設備都被棄置不用，不僅造成先前投入之心力、人力與資源的損失，而且也體現掌權者及許多國人對大型資料庫的認知不足，有待進一步宣導與加強。

針對此項需求，本文特提供下列論述與建言，以增進國人對大型資料庫的正確認知，並供相關教育決策及從業人員參考與討論，以期未來教育資料庫的建置工作能更順暢、設計更完美、內容更完整、運用更廣泛。

二、論述與建言

(一) 大型教育資料庫的特徵與效益

大型資料庫之所以受到重視，主要原因是: 1）資料多元，豐富、完整，2）人數或機構多，涵蓋各群組、具代表性，3）質料信度與效度高、正確可靠，因此可以供研究人員深入分析與探討多種教育議題，是教育評鑑與研究的極佳資源。尤其是具有長期追蹤性質的資料，如國內的 TEPS，美國的全國學生抽樣追蹤研究（National Longitudinal Studies）（https://NCES.ed.gov），以及德國的國家教育追蹤研究計畫（National Educational Panel Study） 因有時序的安排，可以用來模擬前因後果的分析，評估教育政策及方案實施成效，是非常珍貴難得的資料。另外，建置一個大型資料庫供很多研究者使用，

其資料收集單位成本也可能比個人收集總合成本低很多。換言之，大型資料庫的總體效益，要比許多小型獨立、分散收集的資料總和效益大得多，因此值得政府匯集多方資源，建置各教育層級大型資料庫，供教育專業人員使用。

(二) 建置大型教育資料庫應為教育部重要職責之一

由上可知，內容豐富、品質高又便宜的資料，應是教育專業人員共同的期望。不過，建置大型資料庫不是一件容易的事，往往需要政府的支持與協助才能順利達成。因此，若教育部能將建置大型教育資料庫視為重要職責之一，對全國教育評鑑與研究工作的推展將是一項很大的服務與貢獻。

美國教育部所屬的國家教育統計中心（National Center for Education Statistics, NCES），即是一個很好的典範。不僅負責全國性以及一些國際性教育資料的收集，做為制定國家教育方針、政策與措施的依據，而且也積極地釋出大量資料供大學教師與研究生以及地方教育專業人員使用，促進教育研究與評鑑工作，並營造以實證資料為主軸的決策與思判文化。

國家級資料被廣泛使用，可以從美國教育協會（American Educational Research Association, AERA）年會中發表的研究論文看出，很大部分的論文，都是使用 NCES 所收集的國家級或世界級教育資料，分析所得的成果報告。這些論文也因前段所述原因，比較容易得到教育期刊接受刊登。國內學者想要投稿國際期刊，不妨多使用具公信力的大型資料庫的資料來做研究。

(三) 建置大型教育資料庫需集思廣意，細心地規劃

建置一套完整的教育資料系統，除了需要上述政府的支持與輔助之外，還有許多建構要素需要集思廣意，細心地去規劃，包括有關教育政策與措施目前與未來可能發生的議題，應涵蓋的資料內容，最佳資料來源、收集方法、執行程序、資料管控與釋出程序、以及維護與永續經營策略等。以德國國家教育追蹤研究（The German National Edcational Panel Study，NEPS） 方案為例，其規劃過程從界定國人關注的教育議題開始，到資料運用與推廣為止，中間經歷過深思熟慮，集思廣意的項目依序是；多年齡層組追蹤研究方案的選擇與設計（multicohort sequence design），資料內容架構（major data dimensions in different cohorts）、抽樣設計與樣本追蹤方法、資料數據碼設定（coding scheme）、各類資料檔之建置、資料安全維護步驟，資料檔案說明書編制（documentation），資料釋出方式等。每項過程都需要花相當多時間討論後才決定(詳情請參閱 Mauricem， FuB, & RoBbach, 2016）。由此可知，為求周全與完善，建置大型資料庫往往是一項極富挑戰性的問題，不宜堅持個人主觀意見，必須結合相關單位與人員之想法與需求，才能達成目標。

(四) 資料庫建構應先建立法源基礎

為讓資料庫的建置能順利推展，應先建立法源基礎，明定在何種情況或條件之下，政府及學術機構有權力與責任建置與維護資料庫，也明定資料提供者（如學生、家長、教師、校長）參與及合作的義務。例如美國 NCES 單位的設立，即有明文規定其權責。在各級教育法案中也含資料收集與合作條款，讓政府有權向學校及學生等收集所需評鑑、研究與管理等相關資料。以美國高等教育法案為例，其中即明訂凡是接受聯邦政府補助的校院，都必須向 NCES 按時提供所要求呈報的資料，否則將受稽核調查與接受懲處。

當然個人資料隱私權需要受到保護。各國都會有個資法，明訂個資收集與使用規範，保障個人隱私權力，嚴懲非法使用資料去傷害個人，並要求主導機構建置資料安全與正當使用管控機制，確保資料安全與適當使用。

有了法源基礎之後，一切依法行事，會免除許多爭議，讓資料收集與使用工作更順暢，希望國內立法機關能早日建構完整的法源，協助建置大型教育資料庫。

(五) 資料庫建置可從彙整既有紀錄資料開始

資料收集很費人力與經費，所以各學校既有記錄性資料，應優先加以整合善用之。各級學校都有很多學生及家庭背景資料以及教師教學資料，散落各處室，如教務、學務，輔導處等。這些資料是學校裡最完整、正確、客觀的資料，應該好好彙整起來使用。美國國家教育統計中心（NCES）的大學院校的綜合性資料系統（Integrated Postsecondary Education Data System, IPEDS）即是一個好例子。每年各大學校院都得在既定時間內上線呈報多種校務資料到 NCES，經審核通過後彙整成為全國資料庫，供各級政府、學校及一般國民使用。

目前國內大學校院也要呈報教育部基本校務資料。可惜呈報的是一些組合過的資料（processed data），而非原始資料（raw data），沒有辦法重新分析以及和其他個人資料鏈結，進行更細緻的分析，大大地降低資料的分析功能與價值。

建議教育部與學校合作，先規劃全國統一的校務資料定義與檔案規格，供全國學校使用，以利未來資料交換和彙整，建置第一類國家級大型教育資料庫。

第二類大型資料庫是全國性普查或抽樣調查及測驗所得的資料組成。比如學生個人的學習投入程度、校外學習資源與機會等的調查，和學習成果與能力評量，以及從教師以及行政主管（如校長） 和家長做調查所收集到的資料。

以上兩種資料若能鏈結起來，將發揮更大使用價值，條件是校務資料必須是以個人為單位的資料（raw data）。期待不久的將來，能有這種大型資料庫，成為最佳教育研究資源。

(六) 資料內容架構應兼顧政策、學術及使用者之需求

大型資料庫的建置要兼顧政策、學術及其他使用者之需求。為達到此需求，除了舉行聽證會之外，可設置規劃諮詢委員會，徵選各相關領域代表組成，協助設定研究議題與資料內容，以確保兼顧各界需求。一般作法是由委員會選定國家優先議題，其他細節再由執行單位綜合相關單位提供，納入規畫實施。另外，也可以設置技術諮詢委員會，審查抽樣樣本設計與程序以及統計分析方法等，以確保資料品質以及使用適當的測驗及統計分析方法。

(七) 資料庫應視為國家資源，應盡速釋出供國人使用並徵求回饋意見

資料收集最終的目的是使用，使用越多，價值越高，所以必須將資料有效地釋出，以發揮最大價值。資料釋出可採線上分析機制與資料電子檔兩種方式。

線上分析機制是運用特別製作的統計分析軟體，讓資料使用者透過網路直接注入統計分析指示（statistical analysis programing）。軟體會依指示自動進入資料庫，摘取所需資料，完成統計分析，並且只將分析結果列出。另外，任何交叉分析組合（analytical cells），人數少於五人者，資料也將自動消除，完全空白。使用者只知道資料庫裡有哪些資料可作分析，但看不到，也拿不到個人任何資料，所以個人資料不會洩漏，個人資料隱私權受到保護。

以資料電子檔釋出，一般分限制性使用檔和公眾使用檔兩個版本。每版本都會有詳細檔案說明書（coding specifications），列出各資料（variable）的位置（file location）、名稱（variable label）、定義 （short description）和各數據代表的意義（如性別中 1 是男性，2 是女性），以利分析者了解檔案結構、內容與數據意義。

以美國教育部 NCES 為例，公眾使用檔（public-use file），可供一般大眾使用。在此檔中，凡是可能被用來識別個人身分的資料都被移除或組合過（confidential editing），洩漏個人資密的機會已降低到合理的程度，因此任何人都可使用，但還是要求使用者嚴格遵守資料保密規則。

公眾使用檔往往不能滿足某些較深入或更複雜的統計分析，因此另外釋出限制性使用檔（restricted-use file）。此檔的資料沒有經過移除或組合程序（confidential editing）。使用者必須申請並簽署合同（agreement），答應嚴密遵守合同的規定，包括如何安全傳遞與儲藏資料 （含統計時選出的資料），以及合理使用資料、和確保發佈的研究報告的內容，不會用作辨認個人或機關單位之用等。同時，使用者還要接受不定期的訪查。若有違背合同規定，必須退還資料檔，並吊銷其執照。若有洩露資料情事，將受法律嚴厲的處分。最高坐牢五年和罰款二十五萬美元。（詳情請見 http://nces.ed.gov/statprog/instruct.asp。）

為鼓勵使用資料庫，以往 NCES 還辦資料分析研習會，教導學員了解資料庫內容以及合適、正確的統計分析方法與分析結果解讀。目前另有線上教學，讓更多人學習。

由上可知，各國政府都咸認資料庫為國家寶貴資源，應盡量設法讓國人能使用。沒有國家政府會以個資保密理由，拒絕釋出資料。個資保密是應該執行，但總應有方法兼顧保密以及發揮資料的運用價值。

最後建議要求資料庫使用者，在線上依格式提供摘要及建言寄回，倉儲於特設網站上，以利了解資料使用狀況及經驗交流。另外，亦宜共同舉辦年度成果發表會，邀請使用者參與報告，並獎勵研究成品優異者。

(八) 資料庫建置應含永續營運策略

許多資料庫，如追蹤調查（longitudinal studies）和趨勢檢視（trend monitoring），需永續經營，定期進行，其資料價值會更高。再以美國 NCES 為例，許多資料系統，雖有調整更新，但已延續幾十年，甚至上百年，令人敬羨。

要能永續，如前所敘，首先要訂定法源。相關單位得以編制預算，依法執行，不受人事變動或掌權者個人喜好而存廢。一個資料庫可以視社會變遷，修正資料內容規範，但不宜隨意廢棄機制，以免前功盡棄，漠視前人的心力與貢獻。

另外，實務經驗宜保存，至少要保存技術性文件檔案，以利傳承。人員與機構也最好能延續，避免人員重複學習和訓練以及設備的搬遷或重建，耗費資源。

三、結語

以上是一些從實務工作中所得到的經驗與啟示，相信對國內建置大型教育資料庫有所幫助。衷心期待不久的將來，在政府積極引導下，國內很快就會有國際水準的各級教育原始資料（raw data）系統，不僅能透過特製分析系統，快速提供各級教育決策與管理單位所需資料，也能提供教育專業人員豐富的原始資料，進行廣泛教育科研工作，為教育各種議題提供答案，協助國人邁向教育卓越願景，成為國際教育典範大國。

參考文獻

■ Mauricem J., FuB, D., & RoBbach, H. （2015）. Establishing a Comprehensive large-scale data infrastructure for educational research: The example of the German National Educational Panel Study. *Contemporary Educational Research Quarterly, 3* (4), 9 – 40.

■ National Center for Education Statistics, U. S. *Department of Education*, Website address: https//:nces.edu.gov/